# 关于智能编曲的几种神经网络架构

本周我注重阅读了有关智能编曲的几种实现方案，主要是有关智能编曲的神经网络架构，以及关于各种架构的数据利用方法，重点还是在于RNN和LSTM。

先说说Magenta的由来吧，这其实是个挺有意思的事。根据 Quartz 网站的报道，Google 在去年发布的产品 Deep Dream，成为创造 Magenta 的灵感来源。从功能上来说，Deep Dream 可以把原始图像变成迷幻而离奇的超现实风格图片。从性质上来说，Deep Dream 是一款人工智能识图软件，目前已经更新到第 11 版。Google 本意是调教电脑，使之对图片进行描述，但是 Deep Dream 却过度诠释了图片，即选择图片中无关紧要的元素并一再夸大和重复。在 Deep Dream 里，一朵飘动的云可能就会变成一条长相古怪的鱼，或是一只长了若干个头的狗。Magenta 的功能和 Deep Dream 比较相似，即编辑简单的音乐，并且将之转变成辨识度更高的音乐。开发者展示了由 Magenta 制作的一段视频。在视频中，Magenta 依据自己的音乐“知识”基础，通过查找编曲模式和调整曲调，把简单的几个音谱成基本的旋律。或许在短时间内，电脑制作的音乐还不足以冲击流行音乐排行榜。开发者也承认，他制作 Magenta 软件，并不是为了取代人类音乐家。但尽管如此，Magenta 依然象征着计算机在音乐技术上的一大进步，标志着计算机日益重要的地位。编曲是一项艺术工程，它需要艺术家既有天马行空的想象力，也有精湛无双的技术手段，两者缺一不可。我们现在依靠Magenta进行编制的曲子大多还是停留在比较粗疏的阶段，远远谈不上优美动听，可是也正因如此，才需要我们进行不懈努力，让它越来越有“想象力”，也越来技法娴熟。

那么问题来了，如何让它达到我们要有的目标呢，对于人工智能来说，解决方案多种多样，但是归结起来无非是采用更科学高效的神经网络架构和训练方法。其实这样说来现有的人工智能还是存在它的局限性，再怎么改也只是仿照人脑结构，远远达不到和人脑一样可以灵活处理各种问题的地步，而且训练本应是个量变到质变的过程，也就是说人脑可以存在所谓的“顿悟”，然而对于人工智能来说要做到这一点难于上青天。

不过现有的神经网络架构对于编曲来说还是够用的，先把奇思妙想放在一边，主要考虑如何让计算机编曲技术提高一些。对于不同风格不同类型的音乐来说，编曲的神经网络也不尽相同，Eck et al使用两种不同的LSTM（Long-Short Term Memory，是一种时间递归神经网络）网络：一个学习和弦结构和局部音符结构；另一个学习长期依赖，以此来尝试学习旋律并在这个过程中保留它。这使作者可以产生的音乐，却不需要分离原和弦的旋律。然而，这种体系结构训练了一组和弦，却不能创造出更加多样化（新的）的音符组合。

Boulanger Lewandowski等人，试图解决音乐中学习复杂复调结构的挑战。他们使用一个“递归的时间限制Boltzmann机（RTRBM）”，以此建立一个不受现代复调音乐。RTRBM结构的使用使他们可以表现一个复杂的分配对每个时间步长，而不是大多数字符语言模型中一个单一的符号。这使他们能够处理生成音乐中的复调问题。

还有丹尼尔·约翰逊使用RNN实现智能编曲，他提出了一个非常大胆的想法，先看看他要达到的目标：

* 有时间标识：一些理解我想给神经网络参考一次签名当前时间，因为大多数的音乐是由一个固定的时间标识。
* 时不变：我希望网络能够做下去，所以它需要相同的每一个时间步。
* 结构不变：音乐可以自由调换上下，并且在根本上是一致的。因此，我想神经网络的结构对每一个音符来说都是一致的。
* 允许多个音符同时播放，并允许选择相干的和弦。
* 允许同一张被重复：演奏C两次应该不同于双击C。

基于以上种种目标，不变性占据了主流，而卷积神经网络（CNN）就可以在这个方面大展身手了。假设，如果我们用别的东西替换了卷积核，会发生什么?比如说，一个循环的神经网络！每个像素都有它自己的神经网络，它会从一个像素周围的区域接收输入。每一个神经网络都会有自己的记忆细胞和在不同时间之间的周期性连接。

CNN多用于图像，而我们现在使用RNN，用笔记代替像素，我们已经知道我们能做什么了。如果我们让一堆相同的复发性神经网络,一个为每个输出,并给每一个本地的区间(例如,一个八度上方和下方)在注意作为输入,这样我们就有一个不变的系统时间和记录：网络可以在两个方向相对投入工作。

然而，这个网络仍然存在问题。循环的连接允许时间的模式，但是我们没有机制来获得美妙的和弦:每个音符的输出完全独立于每一个音符的输出。在这里，我们可以从上面的RNN-RBM组合中汲取灵感:让我们的网络的第一部分处理时间，让第二部分创造出美妙的和弦。但是，RBM提供了一组输出的一个条件分布，这与每个音符使用一个网络是不兼容的。

最后决定采用的解决方案是，我称之为“双轴RNN”。它的思想是，有两个轴(和一个伪轴):有时间轴和音轴(以及计算的伪轴)。每个循环层都将输入转换为输出，并在其中一个轴上发送重复的连接。但是没有理由他们都必须在同一个轴上发送连接！

注意，前两个层之间有跨时间步骤的连接，但是在notes上是独立的。另一方面，最后两个层之间有联系，但在时间步骤之间是独立的。这使我们能够在不牺牲不变性的情况下，在时间和空间上都有模式。